

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

9-2001

The construct validity of a Belgian assessment centre: A comparison of different models

Filip LIEVENS

Singapore Management University, filiplievens@smu.edu.sg

Etienne VAN KEER

SHL, Belgium

DOI: <https://doi.org/10.1348/096317901167415>

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the [Human Resources Management Commons](#), and the [Organizational Behavior and Theory Commons](#)

Citation

LIEVENS, Filip and VAN KEER, Etienne. The construct validity of a Belgian assessment centre: A comparison of different models. (2001). *Journal of Occupational and Organizational Psychology*. 74, (3), 373-378. Research Collection Lee Kong Chian School Of Business.

Available at: https://ink.library.smu.edu.sg/lkcsb_research/5671

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Short research note

The construct validity of a Belgian assessment centre: A comparison of different models

Filip Lievens*

Ghent University, Belgium

Etienne Van Keer

SHL, Belgium

This study investigates the construct validity of a Belgian assessment centre through a comprehensive set of confirmatory factor analysis models. In particular, the general confirmatory factor analysis approach as well as the correlated uniqueness approach are fitted to the same data. Results replicate Sagie and Magnezy's (1997) finding that the correlated uniqueness model is appropriate to represent exercise effects in assessment centres, as a good fit and no estimation problems are obtained. In this model the dimensions explain 36% of variance, revealing evidence of convergent validity. Possibly, this is due to the careful design of this assessment centre. Evidence of discriminant validity, however, is not established, as the dimension factors are highly correlated.

Since the early 1980s the puzzle of assessment centre construct validity has received increased research attention. Most recent studies have used confirmatory factor analysis (CFA) for testing whether dimensional ratings in assessment centre exercises can be presented by a smaller subset of dimension and/or exercise factors (e.g. Bycio, Alvares, & Hahn, 1987; Kudisch, Ladd, & Dobbins, 1997). Generally, these studies revealed that—when an interpretable solution was found—exercises instead of dimensions emerged as the more valid factors in assessment centres (ACs) (see Lievens, 1998, for a review). In addition, dimension factors were found to correlate rather highly, questioning whether distinct dimensions can actually be measured in ACs. These results are especially puzzling for developmental ACs, which, by definition, require a valid and distinct assessment of dimensions (Bycio *et al.*, 1987; Lievens & Klimoski, 2001).

Recently, Sagie and Magnezy (1997) proposed the correlated uniqueness model (Kenny, 1979; Kenny & Kashy, 1992; Marsh, 1989) as an alternative for the general CFA model to represent AC ratings. This correlated uniqueness model differs from general CFA models in that no separate exercise factors are created. Instead, exercise effects are inferred from the correlations among the error terms of ratings produced by the same exercise. Sagie and Magnezy's study had two important conclusions. First, the correlated uniqueness model yielded interpretable parameter estimates, which implied that the solution reached convergence and that no inadmissible parameter estimates (e.g. factor intercorrelations higher than 1.00 or negative variances) occurred. Second, higher construct validity evidence was found because ratings of psychologist assessors reflected not only the exercises but also the dimensions.

However, Sagie and Magnezy (1997) did not present the fit results of the traditional CFA model so that a comparison between the correlated uniqueness model and the traditional CFA model in the AC domain could not be made. They recommended that 'further research is needed to compare solutions derived from the alternative analytical procedures and to find the most appropriate procedure for the AC setting' (p. 108).

In line with these suggestions, this study sought to replicate Sagie and Magnezy's positive findings with the correlated uniqueness model in another AC and extend them by comparing this model with several traditional CFA models in terms of fit. Accordingly, we evaluate the construct validity of a Belgian AC through a more comprehensive set of models.

These models reflected different conceptualizations of ACs. *Model 1* (Dimensions-only) included only dimension factors. Theoretically, this model was grounded on the traditional trait-based assumption underlying ACs. *Model 2* (Exercises-only) included only exercise factors. In this model the notion of dimensions was abandoned, because an AC was simply regarded as a series of work samples of managerial behaviour. *Model 3* contained one general dimension and exercise factors. This model represented the assumption that assessors were unable to distinguish among dimensions. *Model 4* was a dimensions and exercises model. This model reflected the notion that both dimensions and exercises were inherent AC parts. From a conceptual point of view *Model 5* was similar to *Model 4*. The only exception was that—consistent with Sagie and Magnezy (1997)—this model parameterized exercise effects as correlated uniquenesses instead of as separate exercise factors.

Method

Data originated from an AC designed to identify individuals with potential for higher level management positions. In this AC, 191 managers (149 men, 42 women average age = 37 years; average tenure = 15 years) of a large financial organization were nominated by their supervisors to participate in five exercises: an oral presentation, a role-play with a problem subordinate, two in-baskets, and a past-behaviour interview. The AC targeted six dimensions: activating others, decision making (in crisis), stress tolerance and integrity, effective communication, strategic planning, and commercial skills. Descriptions of the exercises and dimensions are available from the authors. Each AC took place during 2½ days.

Eight trained psychologists (four men, four women; mean assessor experience = 11 years) served as assessors. Assessors were given behavioural checklists to aid their observation, recording and classification tasks (Reilly, Henry, & Smither, 1990). After each exercise, assessors independently provided dimensional ratings (i.e. so-called within-exercise dimension ratings) on 9-point rating scales ranging from 'poor' (1) to 'outstanding' (9). There was one assessor per participant. Consistent with current AC practice, participants were rated by different assessors across exercises. After all exercises, assessors met to discuss their observations and ratings with one another.

All assessors had attended a training seminar in accordance with the Guidelines and Ethical Considerations for Assessment Centre Operations (Task Force on Assessment Center Guidelines, 1989). This training lasted about 4 days. Training content included the explanation of dimensions and exercises used. In addition, the training focused on practice and feedback in the process of observing, recording, classifying, integrating and reporting assessee behaviour. At the end of the training course, the assessors themselves went through the various exercises.

Analysis

We employed EQS (Bentler, 1995) to derive maximum likelihood estimates for the covariance matrix and to test the various models. Similar to previous studies (e.g. Becker & Cote, 1994; Conway, 1996) the appropriateness of the models was based on two sets of criteria. First, a model must not have produced an inadmissible solution such as a failure in the convergence of the iterative estimation procedure or the occurrence of improper estimates. As a second set of criteria for appropriateness we used several goodness-of-fit indices (in addition to χ^2), namely the relative noncentrality index (RNI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). The criteria for evaluating these fit indices were for the TLI and RNI to have values equal to or above .95, and for the RMSEA to be less than or equal to .05.

Results and discussion

Table 1 shows that only a model with six dimensions and exercise effects as correlated uniquenesses (*Model 5*) met our criteria of model appropriateness. In fact *Model 5* had no estimation problems and provided a close representation of the data (TLI and RNI = .99, RMSEA = .032). This result obtained in a Belgian AC replicates Sagie and Magnezy's (1997) findings. As shown in Table 1, the other models indicated either a worse fit or estimation problems. For example, although the CFA model with correlated dimensions and exercises (*Model 4*) provided a very good fit, this model was plagued by serious estimation problems (i.e. four inadmissible estimates: two factor intercorrelations above 1.00 and two negative error variances). This result extends Sagie and Magnezy's (1997) findings, as we found the correlated uniqueness model to perform better than the general CFA models.

Convergent and discriminant validity are further examined by looking at the parameter estimates of the best fitting model. Inspection of the EQS estimates of *Model 5* (a full report is available from the authors) showed that all observed variables had significant loadings on their respective dimension factors and that the mean percentage of variance accounted by dimension factors was 36%. This positive result for convergent validity contrasts with the results of previous studies, which employed the general CFA approach. For instance, in Bycio *et al.* (1987), dimensions accounted for 4% of the variance and in Kudisch *et al.* (1997) for 6%. A possible explanation for the higher dimension variance is that many design factors, which increase the quality of construct measurement (Lievens, 1998), were incorporated in the AC studied. For instance, in this AC fewer dimensions (per exercise), experienced psychologist assessors, and behaviour checklists were used.

Table 1. Summary of goodness-of-fit indices for models

Model	d.f.	χ^2	<i>p</i>	TLI	RNI	RMSEA	90% confidence RMSEA interval	Number of problems
<i>Model 1:</i> 6 Dimensions	75	480.41	.00	0.57	0.69	.171	[.156–.185]	1 ^b
<i>Model 2:</i> 5 Exercises	80	152.51	.00	0.93	0.95	.070	[.053–.086]	0
<i>Model 3:</i> 5 Exercises, 1 general dimension	65	97.36	.01	0.96	0.98	.052	[.028–.072]	0
<i>Model 4:</i> 5 Exercises (correlated) and 6 dimensions	50	48.15	.55	1.00	1.00	.000	[.000–.044]	4 ^a
<i>Model 5:</i> 6 Dimensions and exercises as correlated uniqueness	58	69.02	.15	0.99	0.99	.032	[.000–.058]	0

Note. TLI = Tucker–Lewis Index; RNI = Relative Noncentrality Index; RMSEA = Root mean square error of approximation.

^aTwo factor correlations above 1 and 2 negative error variances; ^bone factor correlation above 1.

Other estimates of *Model 5* (available from the authors) provided less evidence of construct validity. In fact, dimension factors ($M = .73$) were found to correlate strongly, raising doubt on the distinct measurement of the AC dimensions (i.e. lack of discriminant validity). This explains why *Model 3*, including one general dimension and exercise factors, also nearly fitted the data (RMSEA was only slightly above .05, see Table 1). In addition, inspection of the correlated uniquenesses revealed method effects for all exercises (i.e. mean correlated uniquenesses varying from .49 to .70) with the exception of the past-behaviour interview (.18). This positive finding for the past-behaviour interview probably resulted from the structured interview process, which enabled interviewers to obtain multiple behavioural examples per job-related dimension.

The results of this study have implications for both research and practice. In terms of research implications, this study shows that researchers should include the correlated uniqueness model when testing models of AC construct validity. In particular, researchers should adopt the following strategy. First, they should consider which conceptualization (e.g. a model with one dimension and exercises, a model with dimensions and exercises, etc.) is most applicable to their specific AC. Next, they should estimate the CFA model that matches their conceptualization. If the CFA model is plagued by estimation problems, researchers could then estimate the correlated uniqueness model. Besides this theory-driven testing of structural equation models, future research could also examine the applicability of the multiplicative direct product model (Browne, 1984; Campbell & O'Connell, 1967) and hierarchical CFA model (Lance, Teachout, & Donnelly, 1992) to AC ratings. These models have not been tested in the context of ACs.

A limitation of the correlated uniqueness model is the assumption of uncorrelated method effects because correlations are only allowed among the uniquenesses of measures of the same method. If this restriction does not hold, the correlated uniqueness approach suffers from a small biasing effect (i.e. inflation of trait variances and dimension correlations) (Kenny & Kashy, 1992; Lance, Noble, & Scullen, 2000). Hence, researchers planning to use the correlated uniqueness model should strive for methods that are as independent as possible (Becker & Cote, 1994; Conway, 1996). Applied to ACs, researchers are advised not to use the correlated uniqueness model when the AC consists of very similar AC exercises (e.g. three role-plays).

Finally, this study's results provide practitioners with clues on how to improve construct validity. First, we have already noted that the careful design of this AC might have resulted in higher proportions of dimension variance. Therefore, it is important for practitioners to pay attention to such design factors (e.g. using behavioural checklists, experienced psychologist assessors, and measuring only a limited number of dimensions in an exercise) to enhance the quality of construct measurement in ACs. Second, it was striking that method bias was nearly absent in the structured past-behaviour interview. Practitioners might apply this logic to ACs and impose more structure on AC exercises, increasing the opportunities for observing and rating dimension-related candidate behaviours. To this end, role-players trained in eliciting such behaviour from candidates might be used more frequently.

Acknowledgements

Filip Lievens is postdoctoral research fellow of the Fund of Scientific Research, Flanders (Belgium) (FWO). Etienne Van Keer now works at Ernst & Young, Belgium. We would like to acknowledge Michael Brannick, Wilfried De Corte, Martin Evans, and Paul Spector for their suggestions on an earlier version of this manuscript. We are also especially grateful to Paul De Ganck of BBL Belgium. A previous version of this manuscript was presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.

References

- Becker, T. E., & Cote, J. A. (1994). Additive and multiplicative method effects in applied psychological research: An empirical assessment of three methods. *Journal of Management*, 20, 625–641.
- Bentler, P. M. (1995). *EQS: Structural equations program manual*. Encino, CA: Multivariate Software.
- Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices by generalized least squares. *British Journal of Mathematical and Statistical Psychology*, 37, 1–21.
- Bycio, P., Alvares, K. M., & Hahn, J. (1987). Situational specificity in assessment center ratings: A confirmatory factor analysis. *Journal of Applied Psychology*, 72, 463–474.
- Campbell, D. T., & O'Connell, E. J. (1967). Method factors in multitrait-multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research*, 2, 409–426.
- Conway, J. M. (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management*, 22, 139–162.
- Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112, 165–172.
- Kudisch, J. D., Ladd, R. T., & Dobbins, G. H. (1997). New evidence on the construct validity of diagnostic assessment centers: The findings may not be so troubling after all. *Journal of Social Behavior and Personality*, 12, 129–144.
- Lance, C. E., Noble, C. L., & Scullen, S. E. (2000). *The merits of the correlated uniqueness model for multitrait-multimethod data have been oversold*. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.
- Lance, C. E., Teachout, M. S., & Donnelly, T. M. (1992). Specification of the criterion construct space: an application of hierarchical confirmatory factor analysis. *Journal of Applied Psychology*, 77, 437–452.
- Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment*, 6, 141–152.
- Lievens, F., & Klimoski, R. J. (2001). Understanding the assessment center process: Where are we now? In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 16, pp. 245–286). Chichester: Wiley.
- Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, 13, 335–361.
- Reilly, R. R., Henry, S., & Smither, J. W. (1990). An examination of the effects of using behavior checklists on the construct validity of assessment center dimensions. *Personnel Psychology*, 43, 71–84.
- Sagie, A., & Magnezy, R. (1997). Assessor type, number of distinguishable dimension categories, and assessment centre construct validity. *Journal of Occupational and Organizational Psychology*, 70, 103–108.
- Task Force on Assessment Center Guidelines (1989). Guidelines and ethical considerations for assessment center operations. *Public Personnel Management*, 18, 457–470.

Received 2 June 2000; revised version received 3 January 2001